Faculty Working Papers

HOW TO GET THE MOST OUT OF MULTIVARIATE METHODS

Jagdish N. Sheth

#243

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

April 16, 1975

HOW TO GET THE MOST OUT OF MULTIVARIATE METHODS

Jagdish N. Sheth

#243

HOW TO GET THE MOST OUT OF MULTIVARIATE METHODS

Jagdish N. Sheth
University of Illinois
Urbana, Illinois 61801

HOW TO GET THE MOST OUT OF MULTIVARIATE METHODS

Jagdish N. Sheth
University of Illinois

Today multivariate methods are widely used (and often misused) by many marketing professionals and academic researchers. Multivariate methods should or actually have practically replaced the more traditional statistical analyses such as frequency distributions and cross-tabulations in marketing research.

It is, therefore, almost impossible today to come across a research study reported in academic journals such as Journal of Marketing Research which does not utilize some type of multivariate analysis of the data. If multivariate analysis is somehow missing in the study, it is often recognized as a weakness to be rectified in a follow-up research. While this widespread use of multivariate methods has certainly increased the respectability of marketing as a discipline among the more "scientific" and traditional disciplines, it has also brought upon many the problems of communication and understanding. Probably a majority of JMR readers have difficulties comprehending the articles published in it. This problem seems more vivid between the academic and the professional researchers and especially between the researchers and the managers of the marketing function in the organization. Accordingly, there are two objectives of this paper:

1. Provide a nonstatistical description of the multivariate methods and discuss their potential to solve marketing problems.
2. Provide guidelines to the researcher for getting more out of the multivariate methods.

DESCRIPTION AND APPLICATIONS OF MULTIVARIATE METHODS

Multivariate methods refer to those statistical techniques which focus upon, and bring out in bold relief, the structure of simultaneous relationships among three or more phenomena. It is important to remember that what matters in multivariate analysis is the analysis of the simultaneous relationships among phenomena. Many sequential or hierarchial statistical analyses of large number of variables such as the Automatic Interaction Detection (AID) are, therefore, not truly multivariate in nature but simply repeated applications of simpler scatistical techniques.

Multivariate methods differ from simple univariate (single phenomenon) statistical techniques in terms of a shift in focus away from the levels (averages) and distributions (variances) of the phenomena, and instead concentrating upon the degree of relationships (correlations or covariances) among these phenomena. They also differ from the bivariate (two phenomena) statistical techniques by shifting focus away from pairwise relationships to the more complex simultaneous relationships among phenomena [5,6].

Multivariate methods can be broadly categorized into two types: Functional and structural multivariate techniques.

Functional Multivariate Methods

The functional multivariate methods are most appropriate for building predictive models with which the researcher can forecast, or explain one or more phenomena from the knowledge of other phenomena based on their relationships. In order to satisfactorily utilize the functional multivariate methods it is essential that the researcher has considerable knowledge or theory about the market behavior with which to properly conceptualize a realistic

model. The functional multivariate methods provide to the researcher estimates of both the directionality and the magnitude of relationships among phenomena. As such, they border on being the most precise quantitative models of market behavior. Of course, how realistic these models may be is a direct function of the imagination and the experience of the model builder.

Depending on the nature and the number of phenomena the researcher wishes to predict or explain, there are several different types of functional multivariate methods. The first most commonly known and used multivariate method is multiple regression, and its many variations, which enables the researcher to predict the level or the magnitude of a phenomenon such as the sales volume or market share of a brand. The objective in multiple regression is to search for the best possible (optimum) simultaneous relationship between the distribution of the predicted phenomenon and those of the many other correlated or causal phenomena resulting in establishment of a functional relationship between the criterion and the predictor variables. For example, market shares of grocery products may well be a function of a number of marketing mix variables such as average unit price, customer loyalty, media advertising, store coverage, store display and point-of-purchase promotion [1].

A second functional multivariate method is multiple discriminant analysis and many of its variations, which are extremely useful if the researcher is interested in predicting the likelihood of an event happening sometime in the future. For example, what is the likelihood that a tax payer's return will be audited by the IRS? Or what is the likelihood that a telephone customer will convert to TouchTone service when it is promoted by the company? The objective in multiple discriminant analysis is to identify those key discriptors

on which various predefined events have statistically significant differences, and to build a functional model out of them which will enable the researcher to predict likelihoods of events happening as best as possible. Thus, a tax payer with more than $50,000 taxable income or one who belongs to certain occupations such as medical doctors may have significantly higher likelihood of being audited than the average tax payer. Similarly, a household in upper socioeconomic class, younger life cycle or with high mobility may have significantly greater likelihood becoming a TouchTone subscriber than the average customer.

A third functional multivariate method is multivariate analysis of variance (MANOVA) which is more useful for testing the impact of various levels of one or more experimental factors on a variety of phenomena. For example, what is the impact of doubling the advertising budget (weight) on market awareness, attitudes, and purchase behavior toward a product? The objective in MANOVA is to test for significant differences on a set or <u>profile</u> of variables due to some changes in one or more causal factors. Thus, for example, doubling the advertising budget may be highly effective in significantly increasing the awareness and attitude levels among customers but may have little impact on their immediate purchase behavior.

A fourth major functional multivariate method is canonical correlation analysis which enables the researcher to build a predictive model with which he can simultaneously forecast or explain several phenomena based on his knowledge of their correlates. For example, the researcher may be interested in the nature and magnitude of price competition among various brands of a product class. The objective in canonical correlation analysis is to simultaneously regress a set of criterion variables on a set of predictor

variables in the hopes of bringing out the functional relationships both within and between the two sets of variables. Thus it is quite possible that while price competition may be prevalent within the national brands and within the store brands but not between the two types of brands.

Structural Multivariate Methods

The structural multivariate methods, on the other hand, are more descriptive and less predictive in nature. They are essentially data reduction techniques which simplify complex and diverse relationships among phenomena in a manner which enables the researcher to gain insights into the underlying and nonintuitive structure of relationships. The structural multivariate methods are thus analogous to the search for the needle in the hay stack.

The most popular of the structural multivariate methods is factor analysis and many of its variations. Factor analysis enables the researcher to gain insights into the common underlying bonds or dimensions by which otherwise highly divergent phenomena tend to correlate among themselves. For example, what is the common bond between income, education and occupation of a household? Or is there any systematic pattern of preferences in the viewership of vast variety of television programs? The objective in factor analysis is to decompose into meaningful components or dimensions the extent of relationships empirically observed among a set of divergent phenomena. Thus, the common underlying dimension of social class may be responsible for the strong positive correlations found between income, education and occupation. Similarly, interest in situation comedy, quiz shows, soap operas, westerns, police or detective stories, etc. may be the common bondages among the vast variety of television programs.

A second structural multivariate method is cluster analysis which enables the researcher to classify, segment or disaggregate entities into homogeneous subgroups based on their similarities on a profile of information. For example, what are the different psychographic segments of self-medicated drug users; or what are the different benefit segments among bank customers? The objective in cluster analysis is to meaningfully classify a group of entities into clusters based on some judgmental or statistical rule. There are many different algorithms proposed for cluster analysis and very few have any statistical inferential properties so that cluster analysis is more a heuristic than a statistical technique. However, it does provide insights into the typology or segments present in the data. Thus, it is possible to find psychographic segments such as hypochondriacs, skeptics, realists and authority-seekers in the self-medicated drug case [8] , and segments such as social interaction-oriented, banking tasks-oriented and money borrowing-oriented customers in the bank services case.

A more recent structural multivariate method is multidimensional scaling which enables the researcher to explore and infer underlying criteria or dimensions that people utilize to form perceptions about similarities between, and preferences among, various products or services. For example, how do people judge similarities among automobiles or toothpastes? The objective in multidimensional scaling is to map the alternatives in a multidimensional space in such a way that their relative positions in the space reflect the degree of perceived similarity between alternatives. In the process, it provides the researcher insights into the complexity or the number of salient criteria which underlie a person's judgment. Thus, prestige and styling may be the most salient criteria a person uses when he compares various automobiles.

Similarly, decay prevention and brightening of teeth may be the two criteria underlying his judgment about various brands of toothpaste.

Research Needs of Management

The above nonstatistical description of multivariate methods and their applications to marketing research problems clearly suggests that they are highly useful and relevant to marketing. Now let us also look at the potential usefulness of multivariate methods from the perspective of the major types of research needs or inputs for managerial planning. There are four types of research inputs one generally encounters in an organization. They are:

1. Diagonostic research which provides a snapshot representation of the present realities related to products or customers.
2. Prognostic research which trends or forecasts the position in which organization's products or customers are likely to be in sometime in the future.
3. Strategy research by which the organization can assess possible impact of changes in actionable programs on market behavior. This often takes the form of either field experiments or laboratory type simulations
4. Statistical research to ensure that the quality and quantity of information to be analyzed is least biased and most satisfactorily calibrated. The statistical research essentially concerns itself with questions of sampling and nonsampling errors in data, and how to adjust for them by way of analytical strategies.

Table I summarizes the linkage between these four types of research needs and specific multivariate methods relevant to each of those needs. The diagnostic market research is generally exemplified by three areas of research.

TABLE I

Linking Research Needs With Multivariate Methods

Research Needs	Multivariate Methods
A. Diagnostic Research	Structural Methods
1. Market Segmentation	Cluster or Factor Analysis
2. Product or Corporate Typology	Factor or Cluster Analysis
3. Customer Perceptions & Preferences	Multidimensional Scaling or Conjoint Measurement
B. Prognostic Research	Functional Methods
1. Sales Forecasting	Multiple Regression or Canonical Correlation
2. Market Potentials	Multiple Discriminant Analysis
C. Strategy Research	Functional Methods
1. Field Experiments	MANOVA or Discriminant Analysis
2. Laboratory Simulation	MANOVA or Discriminant Analysis
D. Statistical Research	Structural Methods
1. Heterogeneity Reduction	Cluster Analysis or Factor Analysis
2. Measurement Errors	Factor Analysis or Multi-dimensional Scaling
3. Indexing or Data Consistency	Factor Analysis
4. Normal Distributions	Factor Analysis

The first is market segmentation based on some relevant information such as the psychographics,the demographics or the consumption patterns for which both factor analysis and cluster analysis are most appropriate techniques. The second area consists of product, brand or company typology or imagery for which also factor analysis or cluster analysis are useful techniques. The third type of diagnostic market research deals with the why aspect of customer perceptions and preferences about products for which multidimensional scaling techniques are quite relevant.

The prognostic market research is exemplified by at least two types of predictive activities. The first is the forecasting research related to company, industry or product sales either as a time series analysis or as a complex function of environmental and organizational factors. As we discussed earlier, multiple regression and canonical correlation are directly relevant for this area of prognostic research. The second area of research entails estimates of market potentials for new products as well as customer segmentation for existing products. In short, this type of research is directly related to various aspects of the product life cycle. The techniques of multiple discriminant analysis are directly relevant for this type of research.

The strategy market research often entails field experimentation or test marketing. It involves systematic manipulation of marketing mix in selected markets in order to assess their impact on market behavior such as awareness, attitudes and purchase behavior. It is obvious that multivariate analysis of variance is directly relevant here.

The final category is statistical research. There are at least four different aspects of data error and consistency. The first is the question

of heterogeneity. While the present sampling theory is very useful to assist in obtaining a representative sample, there is nothing comparable to ensure a homogeneous sample. On the other hand, a heterogeneous sample has a direct adverse effect on the correlation coefficient which is often reduced to a statistical artifact of aggregating apples and oranges so to speak. The techniques of clustering and factor analysis are, therefore, often used as intermediate stages of analysis to provide insights into the heterogeneity problem. Another area of statistical research is concerned with the question of nonsampling measurement errors inevitable in marketing data. Often it becomes essential to eliminate this error from the data by making appropriate transformations of the data. Once again, factor analysis and multidimensional scaling become very useful intermediate procedures to remove the measurement error from the data. The third area is concerned with the question of consistency of data. Often it is impossible to represent a complex phenomenon such as attitudes or brand loyalty by a single scale. It becomes, therefore, essential to use a variety of indicators which then must be properly indexed to produce a composite score. Once again, factor analysis becomes highly relevant for indexing purposes. Finally, often the raw data does not meet certain statistical assumptions of functional models. This is especially true with respect to the normality assumption. With the use of multivariate methods, it is possible to transform the data so that they are more normally distributed.

HOW TO GET THE MOST OUT OF MULTIVARIATE METHODS

While multivariate methods have direct relevance to marketing problems as we discussed above, it is not easy to successfully implement them in the

research program of the organization due to their novelty, complexity and variety. Therefore, a number of practical guidelines are described below which should be followed by the researcher if he is committed to the idea of integrating multivariate methods in his research program.

First, try not to be technique-oriented. It is not uncommon to find researchers who are comfortable with, and experienced in, a particular multivariate method such as multidimensional scaling factor analysis or multiple regression and try to use that technique across all research problems. They seem to be literally in search for problems which will fit the technique rather than the other way around. Often this leads to redefinition of the problem just so it meets the specifications of the technique. No single technique can solve all research problems, however, and this "Tom Swift and his electronic machine" attitude has resulted in many misapplications of multivariate methods. While it is easy to explain this attitude as due to narrow specializations and discipline biases, it is highly hazardous to the long-term survival of multivariate methods in marketing. In fact, this technique-oriented myopic attitude of the researcher may well become the cause for the downfall of multivariate methods just as it did for operations research models several years ago [6].

Second, consider multivariate models as information inputs to managerial decisions rather than as their substitutes. Often a researcher gets carried away in building models and attempts to replace managerial judgment with the model. Unfortunately this is suicidal in view of the fact that marketing research is only a staff function whose legitimate role is to provide the necessary inputs for managerial decisions. Most managers tend to be satisficers rather than optimizers given the complexity of decisions and

being continuously pressed for time. They regard research as useful input to their judgmental process but do not wish their judgment skills to be replaced by models and computers. In short, it is in the best interest of the researcher to be customer-oriented where his customers are the managers.

Third, multivariate methods or any other technique are not substitutes for researcher's skills and imagination in the proper design of the study. Statistics has nothing to do with causality and can never replace prior theory or experimental design. Unless the problem is adequately conceptualized, it is very easy in today's world of fast,efficient and inexpensive computerized calculations to evoke the GIGO principle (garbage in - gospel out)!

Fourth, half the battle in market research is proper communication of techniques and display of results. It is not at all uncommon to find a brilliant researcher totally competent in multivariate analysis whom the management or even others in the research department simply cannot understand. His communication about the beta weights, heteroscedasticity eigenvalues, varimax rotations, vectors, configurations and Kruskal's Stress are Chinese and Sanskrit to the management. Consequently, the most carefully designed study with highly relevant results for managerial planning go wasted because the management simply cannot understand let alone utilize them as inputs to its decision-making process. It is indeed a sad state of affairs in marketing research that too little emphasis is placed on the art and science of display and communication and too much emphasis is placed on the marginal elegancies of techniques and computer programs.

Fifth, avoid making statistical inferences about the parameters of multivariate models. It is simply impossible in social sciences due to the substantial existence of nonsampling or measurement errors in the data. No

sampling theory can as yet offset this nonsampling error even if one has the resources to sample the total population. Furthermore, it is not easy to apply sampling procedures in social sciences where often we don't know the population itself. Unfortunately, too often multivariate methods have been criticized, chastized and even discarded as irrelevant tools and techniques because it is impossible to make statistical inferences. While it is true that multivariate methods require far more stringent requirements of multivariate normal distributions, it should be pointed out that distribution assumptions underlying statistical techniques even in the univariate and bivariate analysis are also impossible to meet in marketing research.

A better strategy, therefore, is not to discard the techniques as irrelevant but to put them to use for other purposes such as for making substantive inferences or as descriptive statistical techniques by which large data sets can be reduced to meaningful and concise summaries for managerial inputs. In other words, multivariate methods are more useful as data transformation, data reduction and as data display techniques than as mathematical models. This is not the fault of the techniques but the limitations of existing methods of data collection.

Sixth, guard yourself against the danger of making substantive inferences about market realities which may be an artifact solely due to the peculiarities of a particular multivariate method. Since multivariate methods are more complex statistical procedures, there are many more underlying assumptions required for the optimization (minimization or maximization) of statistical decision rules. Consequently, it is easier to inject substantive meanings in the data even if the data are essentially random relationships. This has been especially true of those multivariate methods such as cluster analysis,

multidimensional scaling and conjoint measurement which possess no underlying sampling theory, and therefore, are essentially heuristics often no better than naive judgmental rules.

In order to guard against this danger, it is recommended that the same data be subjected to at least two different techniques. Often, this may be limited to two or more variations of the same basic multivariate method. The replication principle underlying this recommendation will at least bring to the researcher's attention the presence of a technique artifact in his data analysis.

Finally, exploit the complimentary relationship inherent between the structural and the functional multivariate methods. For example, it is extremely advantageous to subject the original predictor variables to a factor analysis and utilizing the transformed factor scores as derived predictor variables in a multiple regression because it makes the data more matching the requirements of lack of multicollinearity and nonsampling error and the presence of normality of the distribution. Similarly, it is best to utilize cluster analysis first to define the number of mutually exclusive groups or segments before attempting a multiple discriminant analysis. In short, this guideline urges the researcher to replace or at least substantiate a number of judgments he has to make in order to build functional multivariate methods, with a structural multivariate analysis of the data. For often the researcher's judgment is highly tenuous and sometimes patently wrong which increases the probability of building less useful multivariate models.

In conclusion, multivariate methods are highly relevant to marketing problems. However, due to lack of familiarity with them, their innate complexity and large variety, it is easy to misapply these techniques. Several

practical suggestions have been made in the paper to increase the likelihood of getting more out of the multivariate methods. Perhaps the single most guideline to recommend is: don't be enamoured by them.

REFERENCES

1. Banks, Seymour. "Some Correlates of Coffee and Cleanser Brand Shares," Journal of Advertising Research, 1 (June 1961), 22-28.

2. Ferber, Robert. Handbook of Marketing Research. New York: McGraw-Hill Book Company, 1974.

3. Gatty, Ronald. "Multivariate Analysis for Marketing Research: An Evaluation," Applied Statistics, 15 (November 1966), 146-58.

4. Sheth, Jagdish N. "Multivariate Analysis in Marketing," Journal of Advertising Research, 10 (February 1970), 29-39.

5. _____, "Multivariate Revolution in Marketing Research," Journal of Marketing, 35 (January 1971), 13-19.

6. _____, "Some Thoughts on the Future of Marketing Models," unpublished Faculty Working Paper No. 232, February 1975, University of Illinois.

7. _____, Multivariate Methods For Marketing Research, Chicago: American Marketing Association (in press).

8. Ziff, Ruth, "Psychographics for Market Segmentation," Journal of Advertising Research, 11 (April 1971), 3-10.